# Supplementary Information

## Supplementary Note

### Mathematical modeling of transcriptional kinetics

#### *Model selection*

Transcription kinetics determines the distribution of nascent mRNA copy number on individual gene loci. Specifically, bursty gene expression often results in a multimodal distribution[1,2]. A widely used model for describing bursty gene expression is the two-state telegraph model[1,3-6], in which the gene randomly switches between an inactive and an active transcription state. If transitions between states are slower than the residence time of nascent mRNA on the transcription site, the nascent mRNA distribution predicted from the model exhibits two Poissonian peaks[6]. Generalizing the model to include more states can create more peaks in the distribution. In general, the number of peaks (i.e., modality) of the experimentally observed distribution sets the minimum number of gene states required for modeling the transcription process.

In this study, the distributions of P1 and P2 nascent mRNA signals both exhibited trimodal distributions (Fig. 5a). One peak in the distribution corresponds to silent loci ($m$ = 0), while two other peaks correspond to two groups of active loci with different expression levels. There are two possible explanations for this phenomenon: (1) individual promoters perform three-state transcription kinetics, and (2) each observed promoter locus is composed of a pair of closely

located sister loci that are indistinguishable under the microscope[7,8]. To evaluate these two explanations, we plotted the distribution of nascent mRNA signals measured from optically resolved sister loci pairs (Fig. 5b). For each promoter, the sister loci exhibited two active populations. The distribution was well fitted by a sum of two Poisson distributions (Considering the intensity threshold used for identifying active transcription sites, the very left part of the distribution (<3 mRNAs) was neglected). By comparing the weights of the two Poisson peaks, we showed that the minor population in the distribution corresponds to 38.4% ± 3.3% of P1 and 17.0% ± 6.1% of P2 sister loci (mean ± s.e.m., data from seven embryos at nc13). Thus, the activity of a single promoter needs to be described with at least three transcription states.

***Model assumptions***

The nascent transcription of each promoter locus was modeled as a three-state process. The model considers three transcription states of the promoter: an “OFF” state (denoted as state 0), where the promoter is transcriptionally inactive, and two “ON” states (denoted as states 1 and 2), where the promoter actively initiates new transcripts. State transitions and mRNA initiations are assumed to be Poisson processes with specific rates $k_{ij}$ and $k_{\mathrm{INI},i}$ ($i$, $j$ = 0, 1, 2), respectively. Following initiation, each nascent mRNA molecule elongates to the final length $L$ with a constant speed $V_{\mathrm{EL}}$. Once completed, the mRNA resides on the gene for an extra termination period, $T_{\mathrm{R}}$, before being released.

At a given observation time, the state of the system at a given observation time $t_{\mathrm{ob}}$ is

determined by the promoter state $n$ ($n$ = 0, 1, 2) and the total signal of nascent mRNA $m$ ($m \geq 0$). Since a nascent mRNA molecule stays on the gene for a fixed period $T_{RES} = L/V_{EL} + T_R$, $m$ is the sum of signals from all transcripts initiated between $t_{ob} - T_{RES}$ and $t_{ob}$, i.e., $m = \sum_{-T_{RES} \leq \tau \leq 0} g(\tau)$. Here, $\tau = t - t_{ob}$ is the time relative to $t_{ob}$. Considering that nascent transcripts may be incomplete, we defined a contribution function $g(\tau)$ to describe the signal from a transcript initiated at time $\tau$ [5, 6]. $g(\tau)$ varies between zero and one, and its exact shape depends on the target positions of the probe set and the magnitude of $T_R$. The shape of $g(\tau)$ for each probe set is shown in Supplementary Fig. 2a.

***Master equation***

We wrote a master equation for the probability distribution of ($n$, $m$) as

$$\frac{d\mathbf{P}(m)}{d\tau} = (\mathbf{K} - \mathbf{K}_{INI})\mathbf{P}(m) + \mathbf{K}_{INI}\mathbf{P}(m - g(\tau)) \quad (1)$$

where $\mathbf{P}(m) = \begin{bmatrix} P(0,m) \\ P(1,m) \\ P(2,m) \end{bmatrix}$ is the probability distribution vector, $\mathbf{K} = \begin{bmatrix} -k_{01} - k_{02} & k_{10} & k_{20} \\ k_{01} & -k_{10} - k_{12} & k_{21} \\ k_{02} & k_{12} & -k_{20} - k_{21} \end{bmatrix}$ and $\mathbf{K}_{INI} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & k_{INI,1} & 0 \\ 0 & 0 & k_{INI,2} \end{bmatrix}$ are matrices describing the promoter-state transition and transcription initiation, respectively[6,9]. Assuming that the marginal distribution of the promoter state at $\tau = -T_{RES}$ is $q(n)$, we can apply an initial condition of $\mathbf{P}(m) = q(n)\delta_{m,0}$ to solve Equation (1) for $\mathbf{P}(m)$ at $\tau = 0$. Specifically, $q(n)$ at steady state satisfies $\mathbf{Kq} = 0$.

The general three-state model allows direct transitions between any two states ($k_{ij} > 0$ for all $i$ and $j$). However, most gene regulation models to date followed the thermodynamic formalism with a detailed balance between states[10]. This constraint limits the topology of the state-transition diagram, i.e., transitions between certain states may be forbidden. Specifically, a three-state model with detailed balance needs to satisfy one of the two schemes of promoter activation, i.e., the sequential activation scheme, in which transitions between states 0 and 2 are forbidden, and the parallel activation scheme, where transitions between states 1 and 2 are not allowed (Supplementary Fig. 5a).

***Mean, variance, and noise***

The mean signal of the nascent mRNA may be derived from Equation (1) as follows[6]:

$$\langle m \rangle = \mathbf{u}\left\{\int_{-T_{RES}}^{0} g(\tau)\mathbf{W}(\tau)d\tau\right\}\mathbf{q} \tag{2}$$

where $\mathbf{u}$ = (1, 1, 1) and $\mathbf{W}(\tau) = e^{-\mathbf{K}\tau}\mathbf{K}_{INI}e^{\mathbf{K}\tau}$. At steady state, the magnitude of $\langle m \rangle$ is proportional to the mean of the contribution function, i.e.,

$$\langle m \rangle = \mathbf{u}\mathbf{K}_{INI}\mathbf{q}\int_{-T_{RES}}^{0} g(\tau)d\tau \tag{3}$$

Thus, the mean nascent mRNA signals measured using different probe sets are in proportion, i.e.

$$\frac{\langle m_1 \rangle}{\langle m_2 \rangle} = \frac{\overline{g_1}}{\overline{g_2}} \tag{4}$$

where $\bar{\cdot} = \frac{1}{T_{RES}} \int_{-T_{RES}}^{0} \cdot \, d\tau$ denotes time averaging. A probe set targeting the 5' region of a transcript should, on average, produce more signal (in units of the number of Bcd molecules) than a probe set targeting the 3' region of the same transcript. In our study, the ratios between the CDS and promoter-specific signals were defined as $a_1$ and $a_2$ in Equation (1). $a_1 < 1$ and $a_2 > 1$ are consistent with the relative target positions of the different probe sets in mRNA sequences.

Moreover, the ratio between different probe signals is quantitatively related to $T_R$, i.e.

$$\frac{\langle m_1 \rangle}{\langle m_2 \rangle} = \frac{\overline{g_{10}} \cdot L + T_R V_{EL}}{\overline{g_{20}} \cdot L + T_R V_{EL}} \tag{5}$$

where $g_{10}$ and $g_{20}$ are the contribution functions for probe sets 1 and 2 with $T_R = 0$. Considering an mRNA elongation speed $V_{EL}$ = 1.5 kb/min[11], we estimated from $a_1$ = 0.53 and $a_2$ = 2.74 that $T_R$ = 142 s for P1 and $T_R$ = 46 s for P2.

Unlike probes targeting the exon or UTR regions of a transcript, the intron probe signal is affected by co-transcriptional splicing (Fig. 2d). Assuming a Poissonian slicing process occurring after the completion of intron synthesis with specific rate $k_{splicing}$, we wrote the average nascent intron signal per locus as

$$\langle m \rangle = \mathbf{u}\mathbf{K}_{INI}\mathbf{q} \int_{-T_{RES}}^{0} g(\tau) s(\tau) d\tau \tag{6}$$

where $s(\tau)$ is the intron survival probability (without being spliced) for a nascent transcript initiated at time $\tau$. $s(\tau)$ is a simple piecewise function satisfying

$$s(\tau)=\begin{cases}1, & \tau > -L_{5'\text{-intron}}/V_{EL}\\ e^{k_{splicing}(\tau+T_{5'\text{-intron}})}, & \tau \le -L_{5'\text{-intron}}/V_{EL}\end{cases} \quad (7)$$

with $L_{5'\text{-intron}}$ denoting the sequence length from the 5' cap to the end of the intron. Thus, the mean nascent P1-intron and 5'UTR signals should be in proportion, with the ratio depending on $k_{splicing}$. The experimental data confirmed this linear relationship and suggested a ratio is of 0.59 (Supplementary Fig. 2e). Assuming that $V_{EL}$ = 1.5 kb/min[11] and $T_{R\text{-}P1}$ = 142 s, we estimated that $k_{splicing}$ = 175 $s^{-1}$. The time scale is similar to that observed in other genes[12-15].

The variance of the nascent mRNA signal was derived from Equation (1) as[6]:

$$\begin{aligned}\sigma_m^2 = {} & \mathbf{u}\cdot\left\{\int_{-T_{RES}}^{0} d\tau_1 g(\tau_1)^2 \mathbf{W}(\tau_1)\right\}\mathbf{q} \\ & +2\mathbf{u}\cdot\left\{\int_{-T_{RES}}^{0} d\tau_1 \int_{-T_{RES}}^{\tau_1} d\tau_2 g(\tau_1)g(\tau_2)\mathbf{K}_{INI}\left(e^{\mathbf{K}(\tau_1-\tau_2)}-\mathbf{qu}\right)\mathbf{K}_{INI}\right\}\mathbf{q}\end{aligned} \quad (8)$$

In case of slow gene-state transitions, $e^{\mathbf{K}(\tau_1-\tau_2)}\approx\mathbf{I}$. Thus,

$$\sigma_m^2 \approx \mathbf{uK}_{INI}\mathbf{q}\int_{-T_{RES}}^{0} g(\tau)^2 d\tau + \mathbf{uK}_{INI}\left(\mathbf{I}-\mathbf{qu}\right)\mathbf{K}_{INI}\mathbf{q}\left(\int_{-T_{RES}}^{0} g(\tau)d\tau\right)^2 \quad (9)$$

Combining Equations (3) and (9), we wrote the noise of the nascent mRNA signal as

$$\eta^2 = \frac{\sigma_m^2}{\langle m\rangle^2} = \frac{1}{\langle m\rangle}\frac{\overline{g^2}}{\bar{g}} + \frac{\mathbf{uK}_{INI}\left(\mathbf{I}-\mathbf{qu}\right)\mathbf{K}_{INI}\mathbf{q}}{\left(\mathbf{uK}_{INI}\mathbf{q}\right)^2} \quad (10)$$

The first term in Equation (10) indicates Poisson noise, which is inversely proportional to $\langle m\rangle$. Its magnitude varies with the shape of *g*. The second term in Equation (10) is due to bursty expression, and the magnitude is invariant with the shape of *g*. For brevity, we rewrote the

expression of noise as

$$\eta^2 = \frac{S_g}{\langle m \rangle} + \eta^2_{\text{burst}} \tag{11}$$

where $s_g = \overline{g^2}/\overline{g}$ is a constant for a given probe set and mRNA species. For P1-specific transcripts, we had $S_{\text{P1-CDS}}$ = 0.78 for the CDS probes and $S_{\text{P1-5'UTR}}$ = 0.99 for the 5'UTR probes.

***Numerically solving the master equation***

Because the analytical solution for Equation (1) is not available, we solved the equation numerically using the finite state projection (FSP) method[9,16,17]. Briefly, we discretized and truncated the range of nascent mRNA signal to $m$ = 0, $\Delta m$, $2\Delta m$, …, $m_{\max}$, with $\Delta m \ll 1$ and $m_{\max}$ large enough to cover the main portion of the nascent mRNA distribution. Equation (1) then transforms to a finite-dimension version:

$$\dot{\bar{\mathbf{P}}} = \left(\bar{\mathbf{K}} + \bar{\mathbf{K}}_{\text{INI}}(\tau)\right)\bar{\mathbf{P}} = \begin{bmatrix} \mathbf{K}-\mathbf{K}_{\text{INI}} & 0 & 0 & 0 \\ 0 & \mathbf{K}-\mathbf{K}_{\text{INI}} & 0 & \cdots \\ 0 & 0 & \mathbf{K}-\mathbf{K}_{\text{INI}} & \cdots \\ \vdots & \vdots & \vdots & \ddots \\ \mathbf{K}_{\text{INI}} & 0 & 0 & \ddots \\ 0 & \mathbf{K}_{\text{INI}} & 0 & \ddots \\ 0 & 0 & \mathbf{K}_{\text{INI}} & \ddots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} \mathbf{P}(0) \\ \mathbf{P}(\Delta m) \\ \mathbf{P}(2\Delta m) \\ \vdots \\ \mathbf{P}(g(\tau)) \\ \mathbf{P}(g(\tau)+\Delta m) \\ \mathbf{P}(g(\tau)+2\Delta m) \\ \vdots \end{bmatrix}, \tag{12}$$

where $\bar{\mathbf{K}} = \begin{bmatrix} \mathbf{K} & 0 & 0 & \cdots \\ 0 & \mathbf{K} & 0 & \cdots \\ 0 & 0 & \mathbf{K} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$, $\bar{\mathbf{K}}_{\mathrm{INI}}(\tau) = \begin{bmatrix} -\mathbf{K}_{\mathrm{INI}} & 0 & 0 & \cdots \\ 0 & -\mathbf{K}_{\mathrm{INI}} & 0 & \cdots \\ \vdots & 0 & -\mathbf{K}_{\mathrm{INI}} & \cdots \\ \mathbf{K}_{\mathrm{INI}} & \vdots & 0 & \cdots \\ 0 & \mathbf{K}_{\mathrm{INI}} & \vdots & \ddots \\ 0 & 0 & \mathbf{K}_{\mathrm{INI}} & \ddots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$.

Next, we discretized the time range $\tau \in [-T_{\mathrm{RES}}, 0]$ into a series with $\Delta\tau \ll T_{\mathrm{RES}}$. The probability distribution of ($n$, $m$) at $\tau = 0$ was computed by propagating the initial state $\bar{\mathbf{P}}_{\tau=-T_{\mathrm{RES}}}$ through the series, i.e.,

$$\bar{\mathbf{P}}_{\tau=0} = \left(\mathbf{I} + \bar{\mathbf{K}}\Delta\tau + \bar{\mathbf{K}}_{\mathrm{INI}}(-\Delta\tau)\Delta\tau\right)\cdots\left(\mathbf{I} + \bar{\mathbf{K}}\Delta\tau + \bar{\mathbf{K}}_{\mathrm{INI}}(-T_{\mathrm{RES}})\Delta\tau\right)\bar{\mathbf{P}}_{\tau=-T_{\mathrm{RES}}}, \quad (13)$$

where $\mathbf{I}$ is the unit matrix. In this paper, we used $\Delta m = 0.1$ and $\Delta\tau = T_{\mathrm{RES}}/2000$ to balance the accuracy and speed of computation.

***Modeling the DNA replication effect***

The fact that some anterior nuclei contain more than two bright FISH spots (Supplementary Fig. 1a) indicates that the *hb* gene in the imaged embryo may have been replicated. Thus, many of the observed bright FISH spots may indeed be a pair of closely located sister loci that are indistinguishable under the microscope[7,8]. To consider this effect in the model/analysis, we note that the two sister gene copies are expressed independently[7,8]. The distribution of the total signal from two closely located sister loci should be a convolution of that of individual ones, i.e.,

$$P(m_{ob}) = P(m_{single}) * P(m_{single}) \tag{14}$$

where $P(m_{ob})$ is the probability of the observed signal from a bright FISH spot composed of two closely located sister loci, and $P(m_{single})$ denotes the nascent mRNA distribution of a single gene copy computed from the model.

In addition to the probability distribution, the low-order statistics of the observed bright FISH spot are also affected by gene replication. Specifically, the mean and variance double with gene replication, while the Fano factor and correlation coefficient stay unchanged.

***Inferring the transcription kinetics***

We fitted the experimental data to estimate the kinetic parameters of each promoter using the maximum likelihood estimation (MLE) method[9,17]. Briefly, we divided the single-locus data of nascent mRNAs from an embryo into multiple subsets according to the nuclear position. To ensure a sufficient number of data points in each subset, we used overlapping binning with a bin size of 0.1 EL. For a given parameter set $\tilde{\mathbf{K}} = \{k_{ij}, k_{INI,i}\}$, the likelihood of observing a subset of data is

$$L(M \mid \tilde{\mathbf{K}}) = \prod_i P(m_i \mid \tilde{\mathbf{K}}) \tag{15}$$

where $P(m_i \mid \tilde{\mathbf{K}})$ is the probability of observing $m_i$ nascent mRNAs given $\tilde{\mathbf{K}}$. For each subset $m_i$, we searched $\tilde{\mathbf{K}}$ to maximize the likelihood in a broad range of parameter values ($k_{ij}$ from 0 to 10 $min^{-1}$, $k_{INI,i}$ from 0 to 100 $min^{-1}$).

To increase the efficiency and robustness of the parameter search for a three-state model, we first fitted a data set pooled from multiple embryos in the same nuclear cycle. For each nuclear position bin, we compared two types of models with either sequential or parallel activation schemes. Using a combination of simplex and simulated annealing methods for the parameter search, we determined that both P1 and P2 data were better fitted by the sequential activation model for all nuclear position bins. Moreover, the results showed that Bcd mainly affected promoter activation rates, while the inactivation and transcription initiation rates remained stable (Supplementary Fig. 5b). Thus, we fixed promoter inactivation and transcription initiation rates at their mean values and re-scanned the activation rates in detail. Once all kinetic rates were determined for the pooled data set, we applied them as initial values to fit the single-embryo data. To increase the accuracy of simplex and simulated annealing methods in the above steps, we repeated each search 12 times. The result with the highest likelihood was chosen.

***Describing P1 and P2 activities using a single model***

Our results showed that P1 and P2 followed common three-state transcription kinetics driven by the same set of Bcd binding events at the two enhancers. Thus, we can combine the description of the two promoters into a single model to relate Bcd binding configurations with P1 and P2 transcription kinetics (Fig. 5i).

The first part of the model describes the Bcd binding dynamics. There are many Bcd binding

sites on the proximal and distal enhancers[18-20]. For simplicity, we assumed that Bcd binding at each enhancer was highly cooperative, with all binding sites being occupied/emptied in one step. This assumption resulted in four possible Bcd binding configurations (Fig. 5i). In the canonical framework of transcription factor binding dynamics, transitions between these binding configurations are described as Poisson processes, whose kinetic rates are related to Bcd concentration by a power law[21]. The steady-state probability of each binding configuration ($s$) satisfies a rational function,

$$P_s(C_{\mathrm{Bcd}}) = \frac{r_s C_{\mathrm{Bcd}}^{n_s}}{\sum_s r_s C_{\mathrm{Bcd}}^{n_s}} \tag{16}$$

where $C_{\mathrm{Bcd}}$ is the Bcd concentration, $n_s$ and $r_s$ are the power-law exponents and proportionality constants for configuration $s$, respectively. Specifically, the configuration with no Bcd bound at either enhancer, (typically denoted as $s = 0$) satisfies $n_0 = 0$ and $r_0 = 1$. For an equilibrium system satisfying detailed balance[10], $n_s$ equals the number of bound Bcd molecules. For a nonequilibrium system, $n_s$ may take higher values[21], yet the general form of Equation (16) still holds.

To relate Bcd binding with the transcriptional activity of a promoter, we assumed that transitions between different promoter states were triggered by specific Bcd binding configurations (Fig. 5i). Bcd binding at a single enhancer (proximal or distal) triggers the transition of a promoter from state 0 to state 1, while the binding at both enhancers triggers the transition from state 1 to state 2. Strictly speaking, these transitions can only happen when the system is at given Bcd binding configurations. However, since transcription factor binding and unbinding

happen at a much faster time scale than promoter activation[8], the promoter activation rates ($k_{01}$ and $k_{12}$) can be modeled as constants over time.

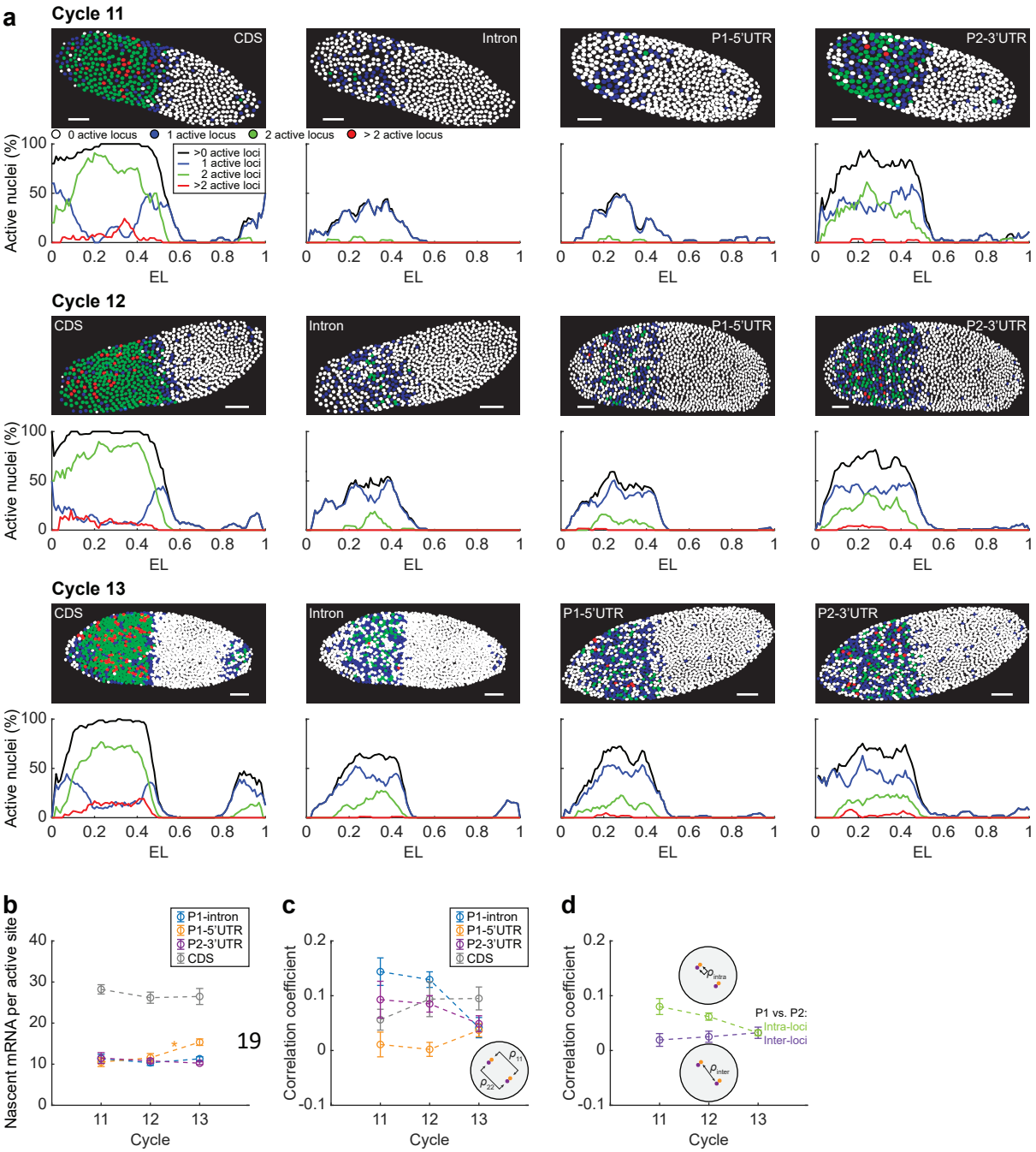
In a simple model, the promoter activation rate may be proportional to the probability of the corresponding Bcd binding configuration. However, activation of a real promoter involves a series of molecular events, some of which are independent of Bcd[8,9]. The activation rate estimated from nascent mRNA distribution ($k_{01}$ or $k_{12}$) represents the overall time scale of all molecular events, i.e.,

$$\begin{cases} k_{01}^{-1} = \left[a_1 P_1(C_{\mathrm{Bcd}}) + a_2 P_2(C_{\mathrm{Bcd}})\right]^{-1} + \tau_{01} \\ k_{12}^{-1} = \left[b P_3(C_{\mathrm{Bcd}})\right]^{-1} + \tau_{12} \end{cases} \tag{17}$$

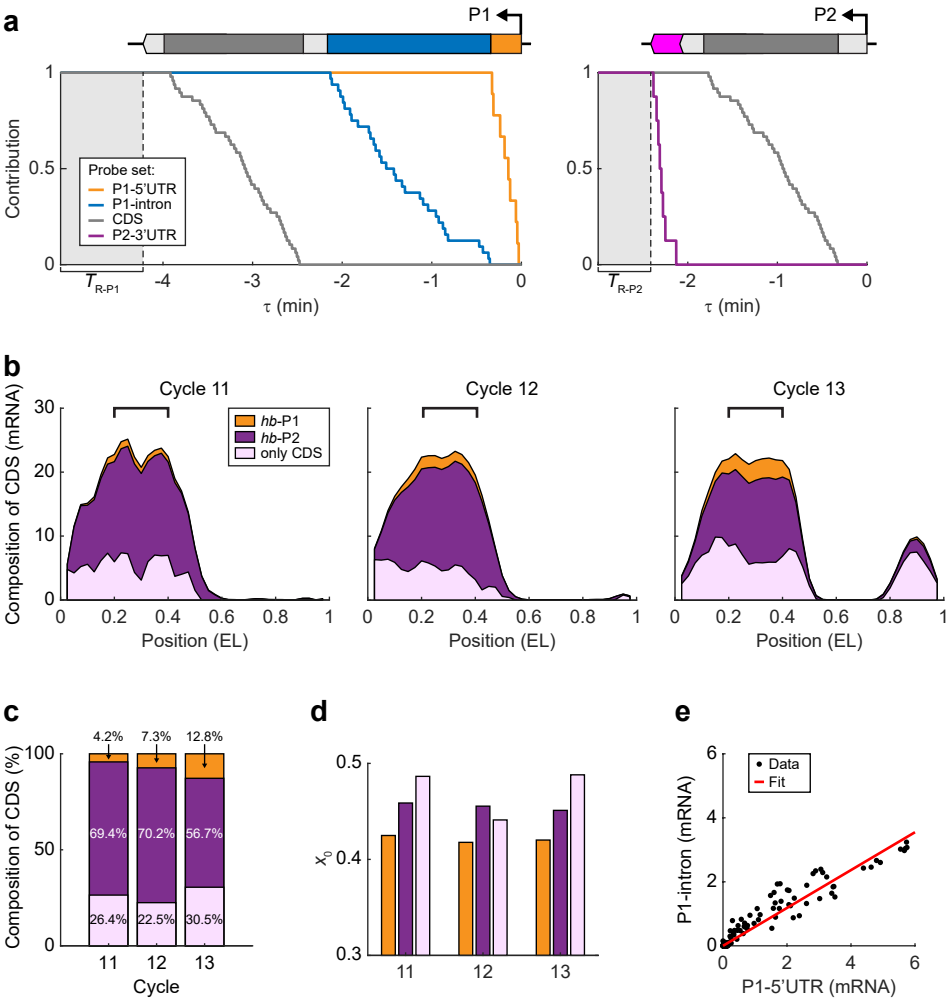
where $a$ and $b$ are proportionality constants and $s$ = 1, 2, 3 denote the binding configurations with Bcd bound at the proximal, distal, or both enhancers, respectively. $\tau$ represents the time scale of Bcd-independent molecular events, which can saturate $k_{01}$ and $k_{12}$ at high Bcd concentration. Equation (17) explains the Hill-function-like relationship between promoter activation rates ($k_{01}$ and $k_{12}$) and Bcd concentration observed in Fig. 5c.

Since P1 and P2 nascent mRNA signals have little correlation (Supplementary Fig. 1d), we speculated that the activation of the two promoters was triggered independently. Thus, the joint distribution of P1 and P2 nascent mRNA signals is the product of their marginal distributions (Fig. 5j).

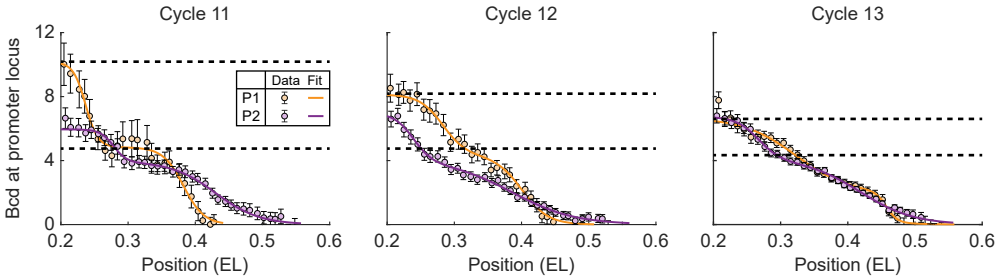
## Supplementary Figures



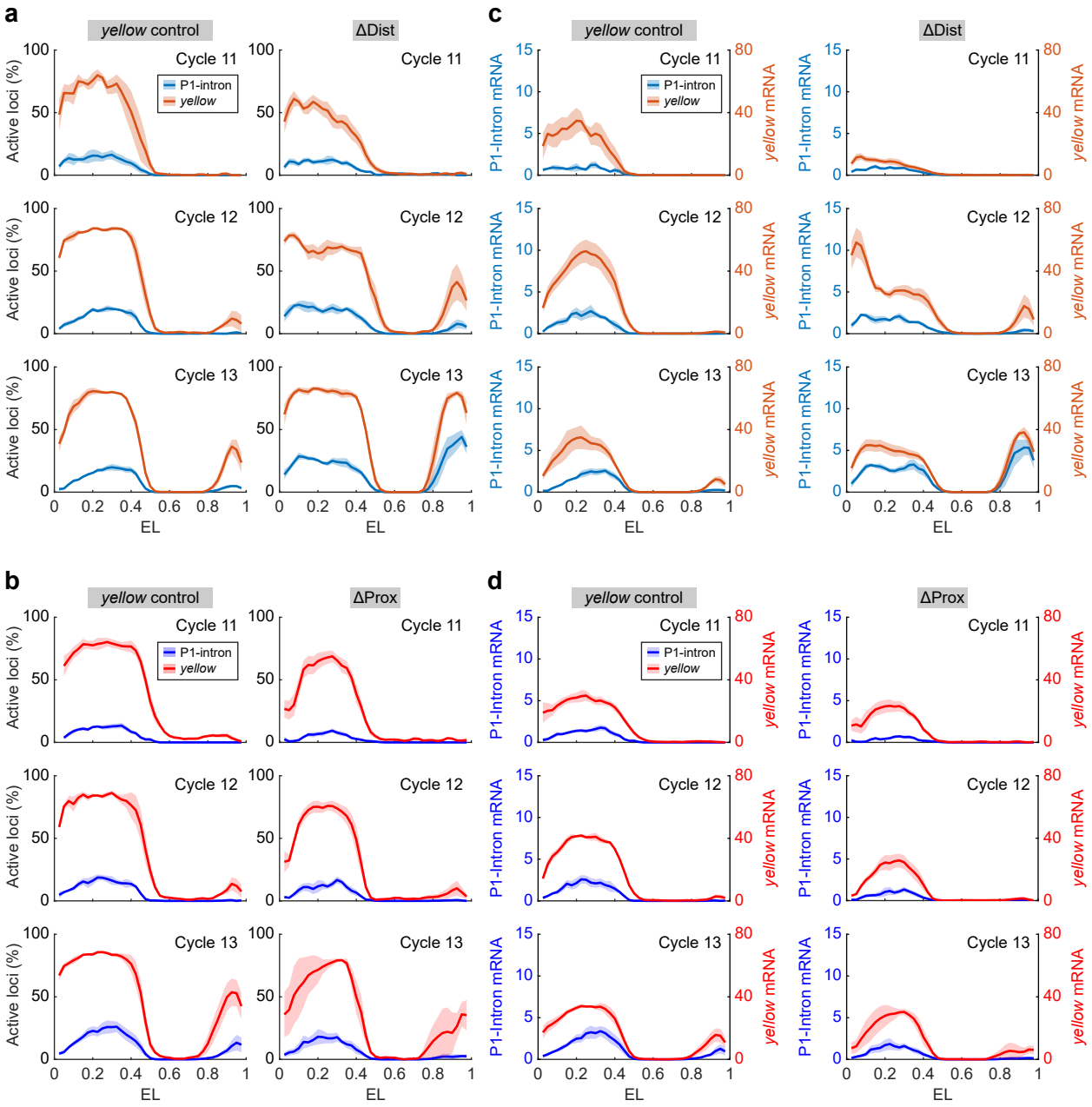

**Supplementary Figure 1. The expression profile and fluctuation of the two *hb* promoters.** (a) The transcriptional activity of individual nuclei in wild-type Drosophila embryos during nc11–13 measured using different probe sets. The color of each nucleus indicates the number of active *hb* loci (see legend). Scale bar, 50 μm. The percentage of nuclei containing different numbers of active loci as a function of the AP position is attached below. (b) The average number of nascent mRNAs per active locus in the position range of 0.2–0.4 EL for different probe signals during nc11–13. Error bars represent s.e.m. (c) The correlation coefficient of the number of nascent mRNAs between different *hb* gene loci in the same nucleus in the position range of 0.2–0.4 EL for different probe signals during nc11–13. (d) The correlation coefficient between P1-5'UTR and P2-3'UTR signals from the same (intra-allele) or different (inter-allele) *hb* gene loci in the same nucleus in the position range of 0.2–0.4 EL during nc11–13. (b-d) Data averaged from ≥5 embryos for each nuclear cycle. Error bars represent s.e.m.



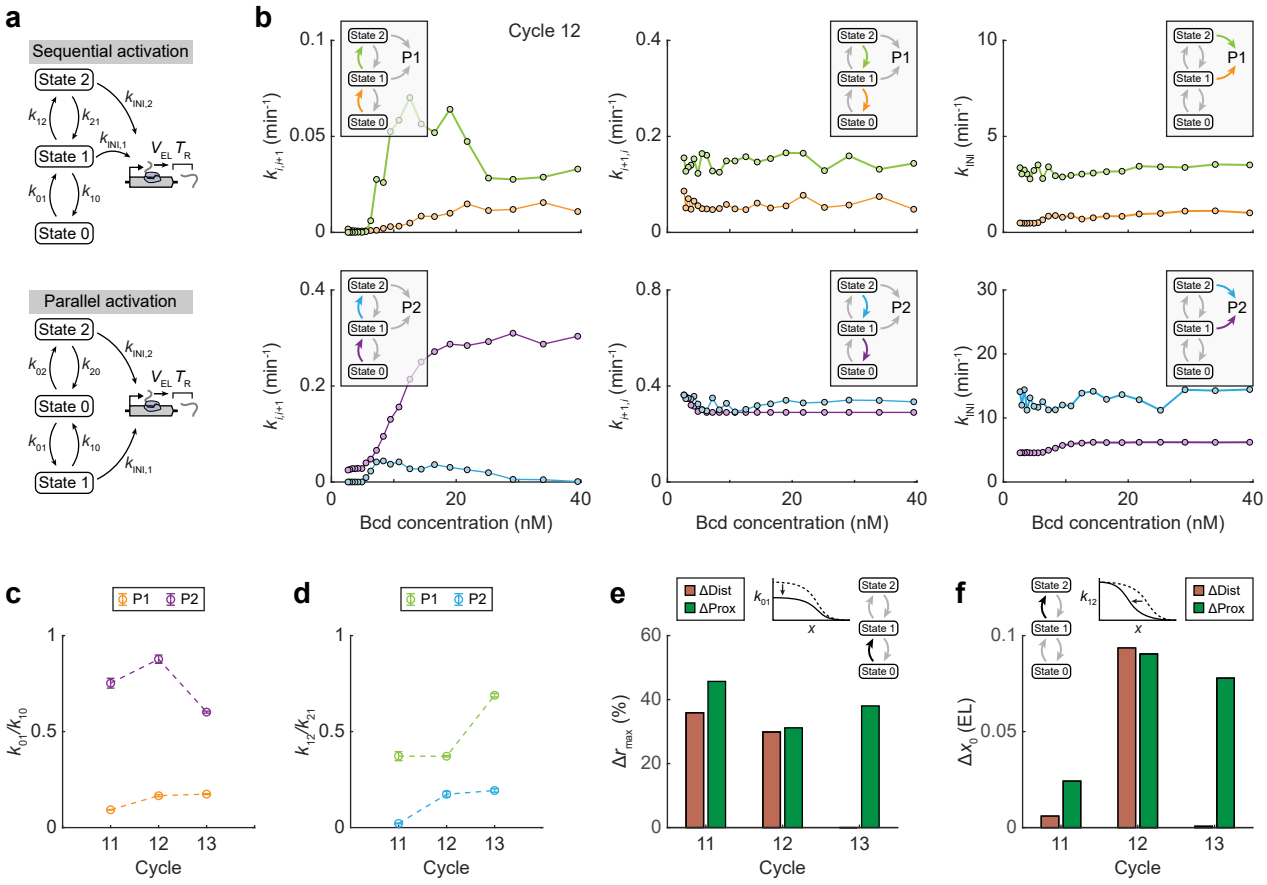

**Supplementary Figure 2. Decomposing nascent *hb* transcription into activities of different promoters.** (a) The contribution functions of different probe sets for P1- and P2-specific transcripts. For each probe set, the observed smFISH signals of a nascent transcript were plotted against the transcript's initiation time. (b) Reconstructing the CDS expression profile from P1- and P2-specific signals for different nuclear cycles using $a_1$ and $a_2$. A substantial fraction of the nascent CDS signal does not correspond to P1-5'UTR and P2-3'UTR signals. Marked region, 0.2–0.4 EL. (c) Average contributions of P1- and P2-specific signals to the nascent CDS signal at individual *hb* loci in the position range of 0.2–0.4 EL during nc11–13. >20% of the nascent CDS signal comes from transcripts not labeled by P1-5'UTR and P2-3'UTR probes. (d) The boundary position of the anterior expression domain for signals with and without promoter-specific labels during nc11–13. (e) The average P1-5'UTR and P1-intron signals per *hb* locus at different AP positions were plotted against each other and fitted to a linear function. Data pooled from 15 embryos during nc11–13 for each probe signal. (b–d) Data averaged from ≥5 embryos for each nuclear cycle.




**Supplementary Figure 3. Quantifying Bcd binding at *hb* locus.** The average number of Bcd molecules bound at P1- and P2-active *hb* loci as a function of nuclear position during nc11–13. For each nuclear cycle, data were averaged from ≥5 embryos and fitted to multi-logistic functions. Dashed lines highlight discrete binding plateaus for each promoter. In the very anterior part of the embryo (<0.25 EL), P2-specific binding curve exhibited an additional plateau with ~6 Bcd molecules. In nc11-12, this plateau is lower than the P1-specific plateau appeared in the same position range. Error bars represent s.e.m.




**Supplementary Figure 4. Comparing promoter activities in different reporter constructs.** (a) Percentage of active reporter gene loci as a function of the AP position for P1-intron or *yellow* signals in the distal-enhancer-removed and control embryos. (b) Percentage of active reporter gene loci as a function of the AP position for P1-intron or *yellow* signals in the proximal-enhancer-removed and control embryos. (c) The number of nascent mRNAs at individual reporter gene loci for P1-intron and *yellow* signals as a function of the AP position in the distal-enhancer-removed and control embryos. (d) The number of nascent mRNAs at individual reporter gene loci for P1-intron and *yellow* signals as a function of the AP position in the proximal-enhancer-removed and control embryos. (a, c) Besides affecting the anterior expression domain, distal enhancer removal also increased the expression of *yellow* and P1 in the terminal regions (0–0.2 and 0.8–1 EL). This observation agrees with a previous report that the distal enhancer may inhibit P1 and P2 at the poles[20]. (a–d) Data averaged from ≥4 embryos for each construct and each nuclear cycle. Shadings indicate s.e.m.




**Supplementary Figure 5. Fitting P1 and P2 nascent mRNA signals using a three-state model.** (a) Schematic of sequential and parallel activation schemes of the three-state model. In the sequential activation scheme, the promoter can be activated from state 0 to state 1 and from state 1 to state 2. In the parallel activation scheme, the promoter can be activated from state 0 to states 1 or 2. (b) The extracted parameters of the three-state model as a function of nuclear Bcd concentration. Data from five embryos at nc12. (c) The average ratio between $k_{01}/k_{10}$ for P1 and P2 in the position range of 0.2–0.4 EL during nc11–13. Error bars represent s.e.m. (d) The average ratio between $k_{12}/k_{21}$ for P1 and P2 in the position range of 0.2–0.4 EL during nc11–13. Error bars represent s.e.m. (e) The relative decrease of the maximal $k_{01}$ level for P1 in the anterior expression domain upon removing one enhancer. In nc11–12, deleting either enhancer caused a decrease of the maximal $k_{01}$ level by ~30%–40%. In contrast, only the removal of the proximal enhancer caused a dramatic decrease of the maximal $k_{01}$ level in nc13, indicating a change in *hb* regulation. (f) The boundary shift of the anterior $k_{12}$ profile towards the anterior pole for P1 upon removing one enhancer. In nc11–12, deleting either enhancer caused an anterior shift of the $k_{12}$ expression boundary. The estimated boundary shift in nc11 is smaller than that in nc12, probably because P1 is much less active in nc11 than in nc12-13 (Fig. 2e, f). In contrast, only the removal of the proximal enhancer caused a dramatic anterior shift of the $k_{12}$ expression boundary in nc13. This is consistent with the decrease of the higher Bcd binding plateau in nc13 (Supplementary Fig. 3) and indicates a change in *hb* regulation. (c, d) Data averaged from ≥5 embryos for each nuclear cycle. (e, f) Data averaged from ≥4 embryos for each reporter construct and each nuclear cycle.

**Supplementary Table 1. Sequences of smFISH probes.**

| Target | Probe sequences (5' to 3') |
|---|---|
| P1-5'UTR | AATGCTGGCGACTTTCGTTT<br>TTTGTATTTCAGTGGCTGCC<br>CGTCCTTTGGATGTTTGGTT<br>CGGGACAAAAGTCTTTTTGC<br>GGATGTGGTCTTTGCCAAAA<br>TTTTGGGCCTCGCTTTTTAG<br>TTAGACCAACACGCACAGTG<br>GGGAGAATTTGGAAACGGGA<br>GGACAGTCCAAGTGCAATTC |
| P1-Intron | TAGGATATTGGATGGTACGC<br>GGAGGTCGAATGCAGTGTAT<br>CTCCCGCAAAAGCGATTTGT<br>GGCGTATGTAAATCCTCACA<br>TTGGATATACGCTGCAGTTC<br>CGTTGCTTAGGAGGAGCAAA<br>AATTAGCTGTCAGGCGTAGA<br>ATTGAAGGGGTTATTTGGGG<br>GTCGAAAATGCATTTGCCGA<br>AGCGTCACATAGGTGTATTC<br>CGGATGACAATCAATTCTGC<br>CCCATTTGTTTGAAATGCGC<br>AAAGCAAGCCAAGAGCAGGA<br>GCAGAAAATGCGCCACACAA<br>CATGTGCTAGTCGTTTTTGG<br>CAGGCAACCGAAACTGCAAA<br>GCCAAACTAAAGGCCAAGTG<br>TGTGTGTGCGCACTATGAAA<br>GCCAAAATTAATTGCTCGGC<br>GCACCACACAAAATGAAGCT<br>GGCATAATTGATGGTTCAGG<br>TTATTATGGGAGGATGGTGC<br>CGGGAAAAAGGGGCATTTAC<br>TGACAACAATTTTCCGCCAG<br>ACGGATCAGAACTGCTTACA<br>AGGATTGCGGGACTTAACTA<br>GGTTTTCTATGGGGATTACG<br>TAGCAGCGAGCTGCGAATTT<br>CGCACTTGGATTTGGATGAT<br>GATCCATTCTGGATTAGAGC<br>CACGCGTCAAGGGATTAGAT<br>TATATCGCTCAGGTAGACGG |
| CDS | TTGTGCTGCTCGTAGTTGGT<br>GAACATGCTGTTGTACCAGG<br>GCTCCTGTTTGATATTTGCC<br>TATTCCCGTCGAGATGATGA<br>AACTGTTCCAGGTGATTGGT<br>ATCCATGGGTTGCTGCTGAA<br>TTTGATCGTTTTGGCTGGGT<br>TTAGCATCGTAATGCTGCAG<br>TTGCTGCAGCAACTGTTGCT<br>TGGAAATGCTGCTGGTACTG<br>ATGGTGATGATGTTGCTGCT |

| | |
|---|---|
| | TTGAATCCACCCATCAGATG<br>TAGAAGTGCTGCATGGGATT<br>TGTTAGTGCCTGCAACTTCT<br>TATTCGACTGACTCGACTTG<br>ATGTACTTCATGTCCTCGCT<br>ATGTTGGTATCATCGTCCTC<br>TGCGAATTGTAGATGGGCAT<br>GGTCTTGCACTTGTAGTTCT<br>TTGTCTGGTTTCATGTGGGT<br>TACTCCAAGTGGTGCTTGAA<br>GTTCTTGTGCTTCCGGATAT<br>ACACGTGTAGCTGCATTTGT<br>GCGAGTTTAGCATGGATTTG<br>TACACAGAACTGTGCGACTT<br>TAATCACAATCCGCACAACG<br>AAGCTGTGGCAATACTTGGT<br>ATACTTGCGCAGATGCAGCT<br>AAACATCGATGACCAACGAG<br>ATTCTTGCTCTTCGGACCAC<br>AGCTGCAACATTTGACTTCC<br>TGGCTGAGATTGCTGTTGCT<br>TTGAACCAGAGGGAATCCTT<br>AAGAAGGCCATGTTGCGGTT<br>TGGAGATTGAGGTTCCAGTA<br>TCGCATTCTTGGCGACAATT<br>TGGTTCTGTTGCTGCAGTTG<br>TGACTTACGCTCGTACTCAT<br>TTCCTTGGGACAGATCCATG<br>TTGTTGCTGCTGCTCATCCT<br>TCCTCCACCTTGAGATTCAT<br>TGTGCTGGGTACTTTCAGTT<br>TTGCTATTGCTGCTGGCATT<br>TTCCATTGCTGCTGGAATTG<br>AGTACTTGCACTCGTAGATG<br>GCGTCCTTGAAGAAGATATC<br>CATGTGAATGGTGTAGAGCA<br>TTGCACTTGAACACATCGTC |
| P2-specific 3'UTR | AGAACTGAGTGTTATGCGCA<br>TCTTCTTTCGTCAGTTTCAG<br>CGCATCTTAGCTACTCTTTA<br>AATTTTGATCCGTTGCTCAG<br>CGACTTAGATTTTATGGGGT<br>GTCTCGAAATTCGTTTCATG<br>ATCAAGGATTACACTGGGCT<br>CATTTCGTGGGCAAATATCT |
| *yellow* | AAACTGCGGTCCATGTTTAT<br>GCCAATCTGGATACGGAATT<br>CAATCTCCAGCTGTATTTGA<br>GTAGGCAGTGGTAATACTGT<br>CCACACTCATCCACTTTAAT<br>ACGGTTCCAGTGTCCAAAAC<br>CACGGATTAGTGGTGGTATT<br>GTATCCGTGGTCAAGTCAAA<br>TAGCTCGTATCTCCGAATTC<br>GTATTTGGATTTGTGTCCAC<br>CACGGCAATGTTAGCTATGA |

| | |
|---|---|
| | TATCCCAATTCATCGGCAAA<br>CCCAGGAGTAAGCAATCAAG<br>AGAATCTCCAGGACTTGTTC<br>CCTCAATGGATCGGGGAAAA<br>CCCATTGGAAGTTAATACCA<br>ATACCAAATATACCCTCCTC<br>AGTACAGGGTACGATAACCA<br>CGATGACTTGCTAACGGACT<br>AAAATCCTCGTGGATACGGC<br>CATGATAGCTATCTTCCGTC<br>CCGTTCATCTAAGGCAACAA<br>ACAGCTCAATTCCATCATCG<br>GAGTACGGCATTGATGAGTG<br>CCACAATGCCATGAAATTGC<br>AACTAAGCCAACGTCATCGC<br>CATCAATTTTCACATCGGCC<br>CCTATCGGATAGAACCCAAA<br>ATCCAAGTCAGACAGCAAGA<br>GGAGCCGTGTAAATTCGGAA<br>AGGCGTTATTCCTCAAATCA<br>TTGAAACGGTATTTGGCGGC<br>GGCAAAACGGCTTGTTTTGG<br>TTCGTATATAACGGTGGACC<br>TTTCTGTGGCAAGACAGGAC<br>CGGGCAAATAAGTGCGACTT<br>TGGAGACTACATTGCCTGAA<br>GGACCCACAGAATTTGTAGA<br>CCGTTGTGCTGGTTGAAAAT<br>GACCACTTGTCTCGTAATTT<br>GGGTTGATGGGTGGGAAATA<br>CGGGCATTCACATAAGTTTT<br>AACCTTGATGCTGATGATGC |

## References


1. Zenklusen D, Larson DR, Singer RH. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat Struct Mol Biol* **15**, 1263-1271 (2008).
2. Munsky B, Neuert G, van Oudenaarden A. Using gene expression noise to understand gene regulation. *Science* **336**, 183-187 (2012).
3. Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol* **4**, e309 (2006).
4. Peccoud J, Ycart B. Markovian Modeling of Gene-Product Synthesis. *Theor Popul Biol* **48**, 222-234 (1995).
5. Senecal A*, et al.* Transcription factors modulate c-Fos transcriptional bursts. *Cell Rep* **8**, 75-83 (2014).
6. Xu H, Skinner SO, Sokac AM, Golding I. Stochastic Kinetics of Nascent RNA. *Phys Rev Lett* **117**, 128101 (2016).
7. Little SC, Tikhonov M, Gregor T. Precise developmental gene expression arises from globally stochastic transcriptional activity. *Cell* **154**, 789-800 (2013).
8. Zoller B, Little SC, Gregor T. Diverse Spatial Expression Patterns Emerge from Unified Kinetics of Transcriptional Bursting. *Cell* **175**, 835-847 (2018).
9. Xu H, Sepúlveda LA, Figard L, Sokac AM, Golding I. Combining protein and mRNA quantification to decipher transcriptional regulation. *Nat Methods* **12**, 739–742 (2015).
10. Mahan BH. Microscopic reversibility and detailed balance. An analysis. *J Chem Educ* **52**, 299 (1975).
11. Garcia HG, Tikhonov M, Lin A, Gregor T. Quantitative imaging of transcription in living Drosophila embryos links polymerase activity to patterning. *Curr Biol* **23**, 2140-2145 (2013).
12. Bentley DL. Coupling mRNA processing with transcription in time and space. *Nat Rev Genet* **15**, 163-175 (2014).
13. Singh J, Padgett RA. Rates of in situ transcription and splicing in large human genes. *Nat Struct Mol Biol* **16**, 1128-1133 (2009).
14. Schmidt U*, et al.* Real-time imaging of cotranscriptional splicing reveals a kinetic model that reduces noise: implications for alternative splicing regulation. *J Cell Biol* **193**, 819-829 (2011).
15. Audibert A, Weil D, Dautry F. In vivo kinetics of mRNA splicing and transport in mammalian cells. *Mol Cell Biol* **22**, 6706-6718 (2002).
16. Munsky B, Khammash M. Transient analysis of stochastic switches and trajectories with applications to gene regulatory networks. *IET Syst Biol* **2**, 323-333 (2008).
17. Neuert G, Munsky B, Tan RZ, Teytelman L, Khammash M, van Oudenaarden A. Systematic identification of signal-activated stochastic gene regulation. *Science* **339**, 584-587 (2013).
18. Driever W, Nusslein-Volhard C. The bicoid protein is a positive regulator of hunchback transcription in the early Drosophila embryo. *Nature* **337**, 138-143 (1989).
19. Struhl G, Struhl K, Macdonald PM. The gradient morphogen bicoid is a concentration-dependent transcriptional activator. *Cell* **57**, 1259-1273 (1989).
20. Perry MW, Boettiger AN, Levine M. Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. *Proc Natl Acad Sci U S A* **108**, 13570-13575 (2011).
21. Estrada J, Wong F, DePace A, Gunawardena J. Information Integration and Energy Expenditure in Gene

Regulation. *Cell* **166**, 234-244 (2016).